Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра информатики

Дисциплина: Машинное обучение (МО)

Отчет о лабораторной работе №9

«Рекомендательные системы»

|  |
| --- |
| **Выполнил:** |
| **Студент гр. 858341** |
| **Немкович А. В.** |

|  |
| --- |
| **Проверил:** |
| **Стержанов М. В.** |

**Минск 2019**

Содержание

[Постановка задачи 3](#_Toc25275643)

[Ход выполнения 4](#_Toc25275644)

[Выводы 13](#_Toc25275645)

# Постановка задачи

Набор данных **ex9\_movies.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит две матрицы Y и R - рейтинг 1682 фильмов среди 943 пользователей. Значение Rij может быть равно 0 или 1 в зависимости от того оценил ли пользователь j фильм i. Матрица Y содержит числа от 1 до 5 - оценки в баллах пользователей, выставленные фильмам.

**Задание.**

1. Загрузите данные **ex9\_movies.mat** из файла.
2. Выберите число признаков фильмов (n) для реализации алгоритма коллаборативной фильтрации.
3. Реализуйте функцию стоимости для алгоритма.
4. Реализуйте функцию вычисления градиентов.
5. При реализации используйте векторизацию для ускорения процесса обучения.
6. Добавьте L2-регуляризацию в модель.
7. Обучите модель с помощью градиентного спуска или других методов оптимизации.
8. Добавьте несколько оценок фильмов от себя. Файл **movie\_ids.txt** содержит индексы каждого из фильмов.
9. Сделайте рекомендации для себя. Совпали ли они с реальностью?
10. Также обучите модель с помощью сингулярного разложения матриц. Отличаются ли полученные результаты?
11. Ответы на вопросы представьте в виде отчета.

# Ход выполнения

**Задание 1**: Загрузка данных ex9\_movies.mat из файла.

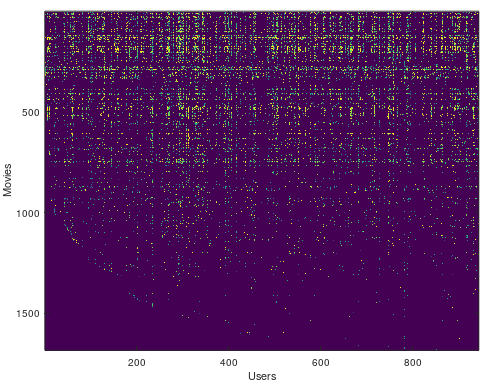


Рис 1. – Демонстрация исходных данных (ex9\_movies.mat)

**Задание 2**: Выбор числа признаков фильмов (n) для реализации алгоритма коллаборативной фильтрации.

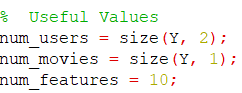
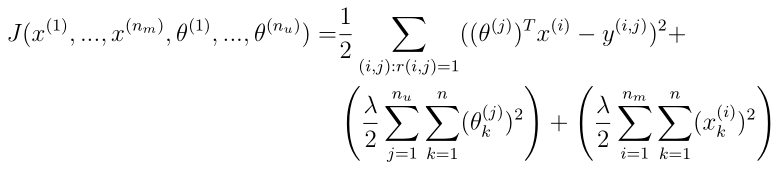


Рис 2. – Код установки числа признаков фильмов (num\_features)

**Задание 3**: Реализация функцию стоимости для алгоритма.

Для расчета функции стоимости для алгоритма коллаборативной фильтрации использовали следующую формулу:



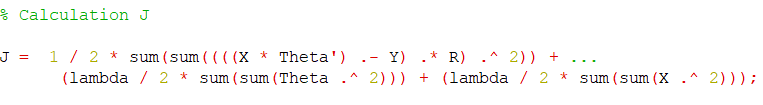
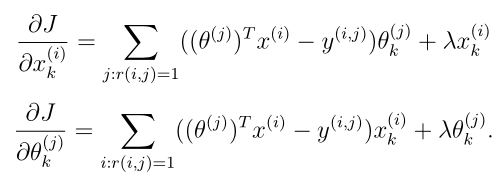


Рис 3. – Код вычисления функции стоимости

**Задание 4-6**: Реализация функции вычисления градиентов и использование векторизации для ускорения процесса обучения, совместно с L2-регуляризацией.

Формулы для расчета градиентов:



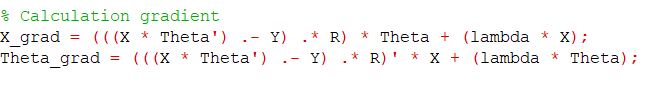


Рис 4. – Код функции вычисления градиентов

**Задание 7**: Обучение модели с помощью градиентного спуска.

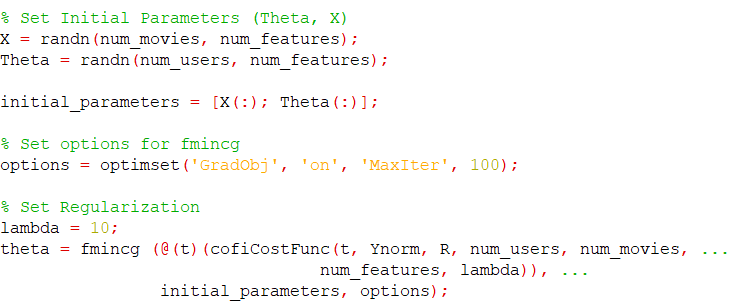


Рис 5. – Код обучения модели с помощью градиентного спуска

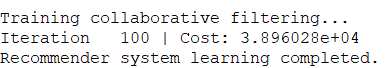


Рис 6. – Обучение модели с помощью градиентного спуска

**Задание 8**: Добавление несколько оценок фильмов от себя. Файл movie\_ids.txt содержит индексы каждого из фильмов.

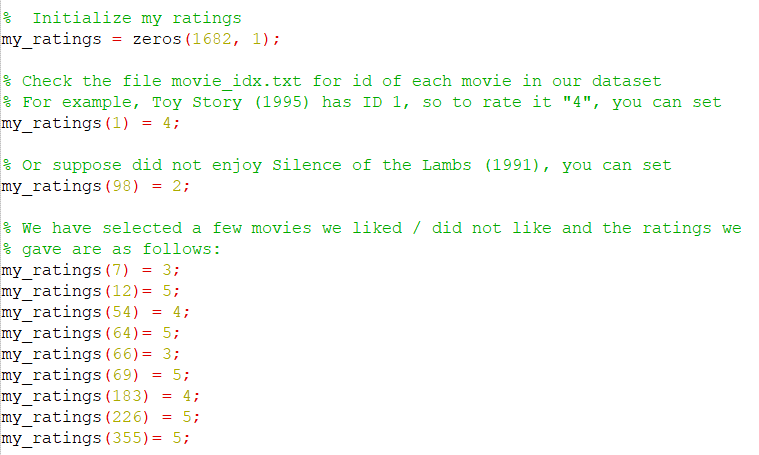


Рис 7. – Код добавления оценок от себя

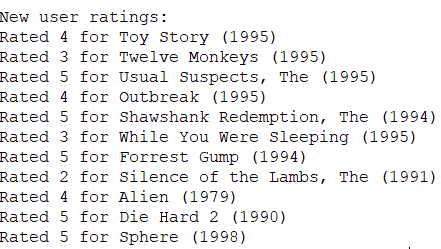


Рис 8. – Оценки от себя

**Задание 9**: Рекомендации для себя.

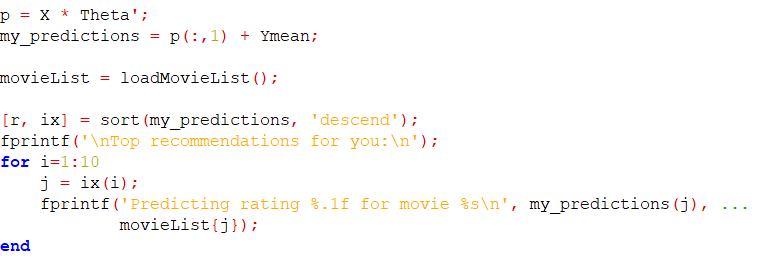


Рис 9. – Код вывода рекомендаций для себя

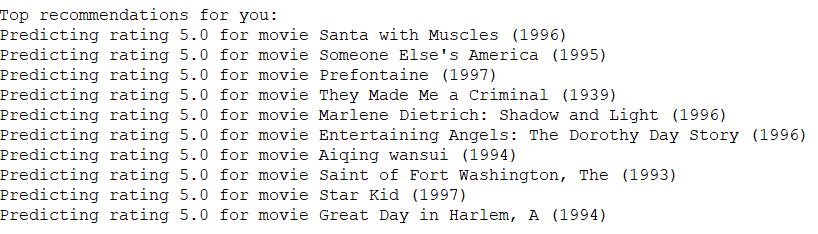


Рис 10. – Рекомендации для себя

# Выводы

Мы научились составлять рекомендации с помощью использования алгоритма коллаборативной фильтрации. В процессе используя градиентный спуск с регуляризацией.

Данный метод помогает предсказать, к примеру оценки фильмов, как для существующей аудитории зрителей, так и для новых.